

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平9-282218

(43)公開日 平成9年(1997)10月31日

(51)Int.Cl. ⁶	識別記号	序内整理番号	F I	技術表示箇所
G 06 F 12/00	5 4 7		G 06 F 12/00	5 4 7 H
	5 1 1			5 1 1
3/14	3 1 0		3/14	3 1 0 B
17/27			15/20	5 5 0 F
17/30			15/419	3 1 0

審査請求 未請求 請求項の数6 O.L (全13頁)

(21)出願番号 特願平8-86989

(22)出願日 平成8年(1996)4月10日

(71)出願人 000004226

日本電信電話株式会社

東京都新宿区西新宿三丁目19番2号

(72)発明者 鈴木 健也

東京都新宿区西新宿3丁目19番2号 日本
電信電話株式会社内

(72)発明者 吉宗 俊哉

東京都新宿区西新宿3丁目19番2号 日本
電信電話株式会社内

(72)発明者 小澤 英昭

東京都新宿区西新宿3丁目19番2号 日本
電信電話株式会社内

(74)代理人 弁理士 志賀 富士弥

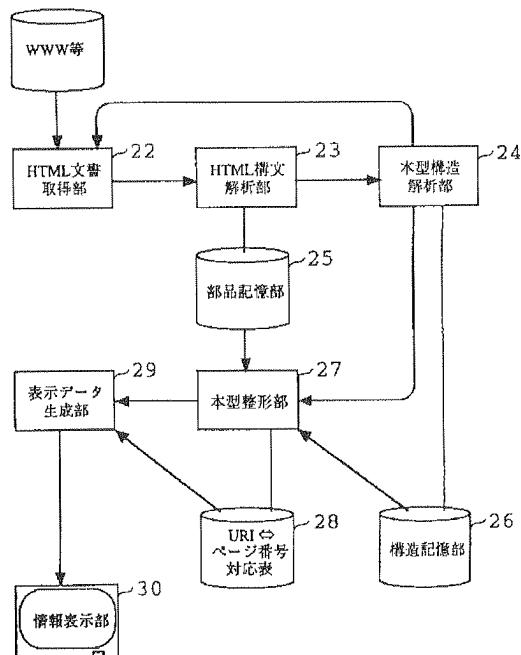
最終頁に続く

(54)【発明の名称】 HTML文書本型整形方法及びその装置

(57)【要約】

【課題】 HTML文書間のリンクに目次、章、節等の論理的構造を記述し、その情報を基にHTML文書を並べ替えて効率的に本の形に整形し表示する。

【解決手段】 HTML文書取得部22はHTML文書を取得し、HTML構文解析部23はそのHTML文書の構文を解析し、HTML文書間の本型の階層や前後関係等の論理的構造の記述としての属性を解釈する。本型構造解析部24は、まずその属性を用いて表現されたHTML文書間の論理的構造を木構造に変換する。次にその木構造を前記属性で表現された文書間の前後関係とできるだけ矛盾のないように並べ替える。最後にこの並べ替えられた木構造を基にHTML文書を線形に並べる。本型整形部27はその並びを基にHTML文書をページへ分割する処理を行い、表示データ生成部29はそれを一ページ毎に表示する形式に変換し、情報表示部30は変換されたHTML文書を本の形で表示する。



【特許請求の範囲】

【請求項1】 インターネット上のハイパーテキスト情報などの情報をタグベースで記述するための構造記述言語であるH T M Lを用いて記述されたH T M L文書を整形する方法であって、任意の情報から他の情報に遷移するための、リンクと呼ばれるH T M L文書内にある識別子に与えられた、複数のH T M L文書間の本型の階層や前後関係といった論理的構造の記述である属性を解釈する第一の過程と、該属性を用いて該論理的構造を木構造に変換する第二の過程と、該木構造を該属性で表現された複数のH T M L文書間の前後関係と矛盾の無いように並べ替える第三の過程と、該並べ替えられた木構造を基にH T M L文書を線形に並べる第四の過程と、から成ることを特徴とするH T M L文書本型整形方法。

【請求項2】 前記第二の過程は、複数のH T M L文書間の本型の論理的構造を記述した目次文書を用意し、該目次文書の記述を用いてH T M L文書間の論理的構造を木構造に変換する過程である、ことを特徴とする請求項1に記載のH T M L文書本型整形方法。

【請求項3】 複数のH T M L文書間の論理的構造が、該リンクの存在する順方向の関係を表現するR E L属性と逆方向の関係を表現するR E V属性とで記述されている場合、この論理的構造を本型の論理的構造の記述に変換する過程を前記第一の過程の前に新たに有する、ことを特徴とする請求項1または請求項2に記載のH T M L文書本型整形方法。

【請求項4】 インターネット上のハイパーテキスト情報などの情報をタグベースで記述するための構造記述言語であるH T M Lを用いて記述されたH T M L文書を整形する装置であって、

H T M Lを用いて記述されたH T M L文書を本型の構造に整形する手段と、本型の構造を画面上に本の形で表示する手段とを備えた装置において、

H T M L文書間の本型の階層や前後関係といった論理的構造の記述である属性を解釈する手段と、

該属性を用いて該論理的構造を木構造に変換する手段と、

該木構造を該属性で表現された複数のH T M L文書間の前後関係と矛盾の無いように並べ替える手段と、該並べ替えられた木構造を基にH T M L文書を線形に並べる手段と、

を備えることを特徴とするH T M L文書本型整形装置。

【請求項5】 前記論理的構造を木構造に変換する手段は、複数のH T M L文書間の本型の論理的構造を記述した目次文書を用意し、これに基づいてH T M L文書間の論理的構造を木構造に変換する手段である、

ことを特徴とする請求項4に記載のH T M L文書本型整

形装置。

【請求項6】 複数のH T M L文書間の論理的構造が、リンクに存在する順方向の関係を表現するR E L属性と逆方向の関係を表現するR E V属性とで記述されている場合、この論理的構造を本型の論理的構造の記述に変換する手段を新たに備えることを特徴とする、請求項4、請求項5に記載のH T M L文書本型整形装置。

【発明の詳細な説明】

【0 0 0 1】

【発明の属する技術分野】 本発明は、インターネットに蓄積されているWWW (W o r l d W i d e W e b) のようなH T M L (Hyper Text Markup Language) 文書を利用者が閲覧しやすい本の形に整形し表示する際に、H T M Lのリンクに本型の論理的構造を記述するための属性を追加し、その属性付きのリンクを用いて本の形に整形するための方法とその装置に関するものである。

【0 0 0 2】

【従来の技術】 従来のH T M L文書を整形し表示するための装置、特にWWWクライアントと呼ばれる装置においては、表示されるH T M L文書を1表示装置につき1文書であった。そのH T M L文書と他のH T M L文書との関係はリンクを用いて表現され、例えそのH T M L文書が他のH T M L文書と一冊の本で表されるような密な関係をもっていたとしても、それぞれは独立に管理される。

【0 0 0 3】 このようなリンクを用いて、H T M L文書間の階層や前後関係などの論理的構造、例えば本のようないくつかの目次、章、節など、を利用者に認識させるためには、「次ページ」、「前ページ」のようなリンクを設定し利用者にそのような遷移を行わせる必要がある。

【0 0 0 4】

【発明が解決しようとする課題】 従来の技術を用いた場合、H T M L文書間に本のようないくつかの目次、章、節などの論理的構造を付与しようとしても、利用者に認識に頼った「次ページ」、「前ページ」のようなリンクを設定する必要があった。また、例え「次ページ」、「前ページ」のようなリンクが設定されていたとしても、それらのリンクは他のリンクと何ら区別されていないために、本の形に整形する際にどのリンクを使って順序づければ良いかという情報が不足し、これを計算機で処理することは難しかった。

【0 0 0 5】 本発明の目的は、H T M L文書間に本のようないくつかの目次、章、節などの論理的構造を記述することができる属性を追加することで、H T M L文書間に本の形に整形し表示することができる方法とその装置を提供することにある。

【0 0 0 6】

【課題を解決するための手段】上記目的を達成するため、請求項1記載の本発明は、インターネット上のハイパーテキスト情報などの情報をタグベースで記述するための構造記述言語であるH T M Lを用いて記述されたH T M L文書を整形する方法であって、任意の情報から他の情報に遷移するための、リンクと呼ばれるH T M L文書内にある識別子に与えられた、複数のH T M L文書間の本型の階層や前後関係といった論理的構造の記述である属性を解釈する第一の過程と、該属性を用いて該論理的構造を木構造に変換する第二の過程と、該木構造を該属性で表現された複数のH T M L文書間の前後関係と矛盾の無いように並べ替える第三の過程と、該並べ替えられた木構造を基にH T M L文書を線形に並べる第四の過程と、成ることを特徴とするH T M L文書本型整形方法であり、H T M L文書を本型に整形することができるこ¹⁰とを最も重要な特徴とする。

【0007】請求項1記載の本発明にあっては、H T M L文書間のリンクに本のような目次、章、節などの論理的構造を記述することができ、従来の技術ではできなかったH T M L文書間の論理的構造を記述することができるようになる。その情報を用いてH T M L文書を並べ替えることで効率的に本の形に整形することができるようになる。

【0008】また、請求項2記載の本発明は、請求項1記載の発明において、前記第二の過程が、複数のH T M L文書間の本型の論理的構造を記述した目次文書を用意し、該目次文書の記述を用いてH T M L文書間の論理的構造を木構造に変換する過程であるとして、H T M L文書間の本型の論理的構造を記述した目次文書のみを与えることで、該目次文書の記述を用いてH T M L文書間の論理的構造を木構造に変換する過程を有するものであり、H T M L文書そのものに本型の論理的構造を記述しなくともH T M L文書を本型に整形することができるこ²⁰とを最も重要な特徴とする。

【0009】請求項2記載の本発明にあっては、H T M L文書そのものは従来の技術で記述されたものでも、H T M L文書間の本型の論理的構造を記述した目次文書を与えるだけで、従来の技術ではできなかったH T M L文書間の論理的構造を記述することができるようになる。その情報を用いてH T M L文書を並べ替えることで効率的に本の形に整形することができるようになる。

【0010】更に、請求項3記載の本発明は、請求項1または2記載の発明において、複数のH T M L文書間の論理的構造が、該リンクの存在する順方向の関係を表現するR E L属性と逆方向の関係を表現するR E V属性とで記述されている場合、この論理的構造を本型の論理的構造の記述に変換する過程を前記第一の過程の前に新たに有し、H T M L文書間の論理的構造をH T M Lのリンクに従来から存在するR E L属性やR E V属性で記述し、該記述を用いて表現されたH T M L文書間の論理的

構造を本型の論理的構造の記述に変換することで、H T M L文書を本型に整形することができることを最も重要な特徴とする。

【0011】請求項3記載の本発明にあっては、H T M Lのリンクの属性に対する拡張は行わなくとも、従来から存在するR E L属性やR E V属性を用いて、階層や前後関係などの本型の論理的関係を記述することで、H T M L文書間の論理的構造を記述することができるようになる。その情報を用いてH T M L文書を並べ替えること¹⁰で効率的に本の形に整形することができるようになる。

【0012】次に、上記目的を達成するため、請求項4記載の本発明は、インターネット上のハイパーテキスト情報などの情報をタグベースで記述するための構造記述言語であるH T M Lを用いて記述されたH T M L文書を整形する装置であって、H T M Lを用いて記述されたH T M L文書を本型の構造に整形する手段と、本型の構造を画面上に本の形で表示する手段とを備えた装置において、H T M L文書間の本型の階層や前後関係といった論理的構造の記述である属性を解釈する手段と、該属性を用いて該論理的構造を木構造に変換する手段と、該木構造を該属性で表現された複数のH T M L文書間の前後関係と矛盾の無いように並べ替える手段と、該並べ替えられた木構造を基にH T M L文書を線形に並べる手段と、を備えることを特徴とするH T M L文書本型整形装置であり、H T M L文書を本型に整形し表示することができるこ²⁰とを最も重要な特徴とする。

【0013】請求項4記載の本発明を用いることで、H T M L文書間のリンクに本のような目次、章、節などの論理的構造を記述することができ、従来の技術ではできなかったH T M L文書間の論理的構造を記述することができるようになる。その情報を用いてH T M L文書を並べ替えることで効率的に本の形に整形し表示することができるようになる。

【0014】また、請求項5記載の本発明は、請求項4記載の発明において、前記論理的構造を木構造に変換する手段が、複数のH T M L文書間の本型の論理的構造を記述した目次文書を用意し、これに基づいてH T M L文書間の論理的構造を木構造に変換する手段であるとして、H T M L文書間の本型の論理的構造を記述した目次文書のみを与えることで、H T M L文書そのものに本型の論理的構造を記述しなくてもH T M L文書を本型に整形することができるこ³⁰とを最も重要な特徴とする。

【0015】請求項5記載の本発明にあっては、H T M L文書そのものは従来の技術で記述されたものでも、H T M L文書間の本型の論理的構造を記述した目次文書を与えるだけで、従来の技術ではできなかったH T M L文書間の論理的構造を記述することができるようになる。その情報を用いてH T M L文書を並べ替えることで効率的に本の形に整形することができるようになる。

【0016】更に、請求項6記載の本発明は、請求項

4、または5記載の発明において、複数のH T M L文書間の論理的構造が、リンクに存在する順方向の関係を表現するR E L属性と逆方向の関係を表現するR E V属性とで記述されている場合、この論理的構造を本型の論理的構造の記述に変換する手段を新たに備えることで、H T M L文書を本型に整形することができることを最も主要な特徴とする。

【0017】請求項6記載の本発明においては、H T M Lのリンクの属性に対する拡張は行わなくとも、従来から存在するR E L属性やR E V属性を用いて、階層や前後関係などの本型の論理的関係を記述することで、H T M L文書間の論理的構造を記述することができるようになる。その情報を用いてH T M L文書を並べ替えることで効率的に本の形に整形することができるようになる。

【0018】以下に、本発明の作用を述べる。

【0019】請求項1記載の本発明において、リンクにH T M L文書間の本型の階層や前後関係などの論理的構造の記述として与えられた属性を解釈する過程は、従来の技術では解釈することができなかった本型の論理的構造を解釈することができるようしている。また、該属性を用いて表現されたH T M L文書間の論理的構造を木構造に変換する過程は、各H T M L文書に分散した論理的構造の記述を一つの木構造として表現することで集中化して扱うことができるようしている。更に、該木構造を前記属性で表現された文書間の前後関係とできるだけ矛盾のないように並べ替える過程は、より上位の階層で記述された前後関係を補助としてH T M L文書間のリンクに記述されている前後関係の順に並べ替えることで、H T M L文書間に前後関係が記述されていない場合や矛盾した記述を含む場合にも正常に並べ替えが行われるようしている。一方、該並べ替えられた木構造を基にH T M L文書を線形に並べる過程は、H T M L文書に記述された本型の論理的構造にできるだけ適合させた木構造を深さ優先で探索することで、H T M L文書を線形に並べている。従って、H T M L文書間の論理的構造を記述し、その情報を用いてH T M L文書を並べ替えることが可能となり、本発明の目的であるH T M L文書を効率的に本の形に整形することができるようになる。

【0020】請求項2記載の本発明において、目次文書の記述を用いて表現されたH T M L文書間の論理的構造を木構造に変換する過程は、与えられた目次文書の記述を文書の先頭から順に展開することでH T M L文書間の階層や前後関係などを得、その情報を用いてH T M L文書間の論理構造を木構造に変換することを行っている。従って、H T M L文書そのものには本型の論理的構造が記述されていなくても、目次文書の記述からH T M L文書間の論理構造を木構造に変換することが可能となり、本発明の目的であるH T M L文書を効率的に木の形に整形することができるようになる。

【0021】請求項3記載の本発明において、H T M L

のリンクに従来から存在するR E L属性やR E V属性の記述を用いて表現されたH T M L文書間の論理的構造を本型の論理的構造の記述に変換する過程は、リンクのR E L属性やR E V属性によって表現された親子関係や前後関係を本型の階層関係や前後関係に変換することを行っている。従って、H T M L文書間の論理的構造を記述し、その情報を用いてH T M L文書を並べ替えることが可能となり、本発明の目的であるH T M L文書を効率的に本の形に整形することができるようになる。

【0022】請求項4記載の本発明において、リンクにH T M L文書間の本型の階層や前後関係などの論理的構造の記述として与えられた属性を解釈する手段は、従来の技術では解釈することができなかった本型の論理的構造を解釈することができるようしている。次に、該属性を用いて表現されたH T M L文書間の論理的構造を木構造に変換する手段は、各H T M L文書に分散した論理的構造の記述を一つの木構造として表現することで集中化して扱うことができるようしている。更に、該木構造を前記属性で表現された文書間の前後関係とできるだけ矛盾のないように並べ替える手段は、より上位の階層で記述された前後関係を補助としてH T M L文書間のリンクに記述されている前後関係の順に並べ替えることで、H T M L文書間に前後関係が記述されていない場合や矛盾した記述を含む場合にも正常に並べ替えが行われるようしている。最後に、該並べ替えられた木構造を基にH T M L文書を線形に並べる手段は、H T M L文書に記述された本型の論理構造にできるだけ適合させた木構造を深さ優先で探索することで、H T M L文書を線形に並べている。従って、H T M L文書間の論理的構造を記述し、その情報を用いてH T M L文書を並べ替えることが可能となり、本発明の目的であるH T M L文書を効率的に本の形に整形し表示することができる装置を提供することができるようになる。

【0023】請求項5記載の本発明において、目次文書の記述を用いて表現されたH T M L文書間の論理的構造を木構造に変換する手段は、与えられた目次文書の記述を文書の先頭から順に展開することでH T M L文書間の階層や前後関係などを得、その情報を用いてH T M L文書間の論理構造を木構造に変換することを行っている。従って、H T M L文書そのものには本型の論理的構造が記述されていなくても、目次文書の記述からH T M L文書間の論理構造を木構造に変換することが可能となり、本発明の目的であるH T M L文書を効率的に木の形に整形することができるようになる。

【0024】請求項6記載の本発明において、H T M Lのリンクに従来から存在するR E L属性やR E V属性で記述されたH T M L文書間の論理的構造を本型の論理的構造の記述に変換する手段は、リンクのR E L属性やR E V属性によって表現された親子関係や前後関係を本型の階層関係や前後関係に変換することを行っている。従

って、HTML文書間の論理的構造を記述し、その情報を使ってHTML文書を並べ替えることが可能となり、本発明の目的であるHTML文書を効率的に本の形に整形することができるようになる。

【0025】

【発明の実施の形態】以下、図面を用いて本発明の実施形態例について説明する。

【0026】〔実施形態例1〕図1は、本発明の第一の実施形態例によって整形された結果の本型の論理的構造をモデル化した図である。同図に示す本1は論理的構造としての本全体を表す。このような本は、通常、まえがき2や目次3、本文4、参考文献目録5、索引6、その他7から構成される。ここで、本文4は更に章8が繰り返されたもので構成され、章8は節9が繰り返されたもの、節9はページ10が繰り返されたもの、ページ11は単語11が繰り返されたもので構成される。まえがき2やその他7も本文4と同様に、節9や章8が繰り返されたもので構成される。その他7には、付録や補追と呼ばれるものが該当する。また、目次3は主に章8は節9のような本内部への参照で構成される。索引6も同様にページ10などの本内部への参照で構成される。一方、参考文献目録5は別の本など本の外部の情報への参照で構成され、本内部の単語11などから参照される。本発明では、ページ10と単語11の論理的構造以外を記述し、ページ10の繰り返しの構造は自動的に作成する。

【0027】図2は、本発明の第一の実施形態例における上記の論理的構造をリンクに記述するために必要となる属性を示した図である。同図に示す属性book12は図1における本1に対応し、本全体を記述したHTML文書からのリンクであることを表している。また、属性section13は図1におけるまえがき2やその他7、本文4の章8や節9に対応し、章や節などのページをまとめた構造のHTML文書からのリンクであることを表している。属性index14は図1における目次3や索引6に対応し、本内部への参照をまとめたHTML文書からのリンクであることを表している。属性bibliography15は図1における参考文献目録5に対応し、本の外部への参照をまとめたHTML文書からのリンクであることを表している。

【0028】属性book12の値としては、まえがき2や本文4、その他7などの属性section13で表されるHTML文書へのリンクの場合は“section”、目次3や索引6などの属性index14で表されるHTML文書へのリンクの場合は“index”、参考文献目録5などの属性bibliography15で表されるHTML文書へのリンクの場合は“bibliography”を与える。また、続巻物小説のように何冊かの本で一つのまとまりとなる本を記述するために、論理的に前を表す“previous”や、論理的に後ろを表す“next”を与えることもで

10

20

30

40

50

きる。更に、作者を表す“made”を与えることができる。次に、属性section13の値としては、リンクの書かれているHTML文書が章ならば節、節ならば項を表すHTML文書へのリンクに対して“section”を与える。また、章や節などの論理的な前後関係を記述するために、前を表す“previous”や、後ろを表す“next”を与えることができる。属性index14の値としては、本内部の情報に対する参照のリンクの場合は“refer”を与える。また、目次や索引などが複数のHTML文書に渡って記述されている場合には、論理的に前を表す“previous”や、論理的に後ろを表す“next”を与えることができる。属性bibliography15の値としては、本外部の情報に対する参照のリンクの場合に、“refer”を与える。また、参考文献目録などが複数のHTML文書に渡って記述されている場合には、論理的に前を表す“previous”や、論理的に後ろを表す“next”を与えることができる。

【0029】本発明の第一の実施形態例における論理的構造を表す上記の属性をリンクに記述した例を以下に示す。

【0030】book.html～16

```
<head>
<link book="made" href="mailto:kenya@ntt.jp">
</head>
<body>
<A book="index" href="mokujii.html">目次</A><p>
<A book="section" href="chap1.html">第一章</A><p>
<A book="bibliography" href="bibl.html">参考文献</A><p>
</body>mokujii.html～17
<body>
<h1>目次<h1><p>
<A index="refer" href="chap1.html">第一章</A><BR>
<A index="refer" href="sec1.html">第一節</A><BR>
<A index="refer" href="sec2.html">第二節</A>
<p>
<A index="refer" href="chap2.html">第二章</A>
</body>chap1.html～18
<body>
<h1>第一章<h1><p>
<A section="section" href="sec1.html">第一節</A><p>
```

```

<A section="section" href=
"sec2.html">第二節</A><p>
</body>sec1.html~19
<body>
<h2>第一節</h2><p>
これは本型表示のテストページです。<p>
<A section="next" href="se
c2.html">次節</A><p>
</body>sec2.html~20
<body>
<h2>第二節</h2><p>
これは本型表示のテストページの第二節です。<p>
<A section="previous" href
="sec1.html">前節</A><p>
</body>bibl.html~21
<body>
<h1>参考文献目録</h1><p>
[1]
<A bibliography="refer" hr
ef="http://www.ntt.jp/">N
TT Home Page</A><BR>
[2]
<A bibliography="refer" hr
ef="http://hil.ntt.jp/">N
TT Human Interface lab</A
><BR>
</body>上記に示すbook.htmlは、
図1における本1に相当するHTML文書で本全体を表
している。<head>タグと</head>タグで囲
まれたヘッダに、属性book="made"をもつ<
link>タグによってこの本の作者を表すリンクが記
述されている。ここで、href="文字列"は、その
文字列をリンクの識別子とすることを示し、その文字列
のことをUniformResource Identifier略してURIと呼ぶ。また、<a>タグによ
って目次や第一章、参考文献目録へのリンクが記述され
ている。mokujii.html17は、図1における
目次3に相当するHTML文書で、book.html
16から属性book="index"のリンクで参照
されている。この文書17には、本全体の構成を表す目
次で属性index="refer"をもつ<a>タグ
によって第一章、第一章第一節、第一章第二節、第二章
への本内部参照を表すリンクが記述されている。このm
okujii.html17は、その他のHTML文書の
内容から自動的に生成することができる。chap1.
html18は、図1における章8に相当するHTML
文書で、book.html16から属性book=
"section"を持つリンクで参照されている。こ
の文書18には、属性section="sectio
n"をもつ<a>タグによって第一章を構成する第一

```

節、第二節へのリンクが記述されている。sec1.html19やsec2.html20は、図1における節9に相当するHTML文書で、chap1.html18から属性section="section"を持つリンクで参照されている。これらの文書19、20には、第一章を構成する第一節と第二節の内容と、それらの論理的前後関係を示すリンクが属性section="next"や属性section="previous"をもつ<a>タグによって記述されている。bibl.html21は、図1における参考文献目録5に相当するHTML文書で、book.html16から属性book="bibliography"を持つリンクで参照されている。この文書21には、この本の参考文献のリストを属性bibliography="hrf"をもつ<a>タグによって記述している。

【0031】図3は、上記に示したHTML文書間の関係を図に表したものである。角丸四角形はHTML文書内でリンクの記述されている部分を示し、矢印がリンクの参照先を示している。ここで、網掛けした角丸四角形は本の外部への参照を表し、矢印に付加された文字はリンクのbook属性やsection属性、index属性の値を表している。

【0032】図4は、上記に示したHTML文書を本発明の第一の実施形態例によって本型に整形した例を示した図である。それぞれのHTML文書は記述された本型の論理的構造に沿って、図1で示したモデルの本1、目次3、本文4、参考文献目録5の順に並べ替えがなされている。本実施形態例では、一つのHTML文書が一ページに収まりきらない場合は、二ページ以上に分割する。図4に示した各四角形は本型整形後の一ページを表しており、下部にハイフン(ー)で囲んだ数字はそのページのページ番号を表している。

【0033】図5は、本発明の第一の実施形態例に係るHTML文書本型整形装置の構成を表すブロック図である。同図に示すHTML文書取得部22は、WWW等のHTML文書を蓄積しているデータベースよりHTML文書を取得しHTML構文解析部23に渡す役割をもつ。HTML構文解析部23では、HTML文書取得部22より渡されたHTML文書の構文を解析し、処理中のHTML文書のURIと本実施形態例で定めた属性をもつリンクを本型構造解析部24へ渡し、処理中のHTML文書を部品記憶部25へ格納する。本型構造解析部24は、本発明の最も重要な部分であり、本型の論理的構造を解析しHTML文書の並べ替えを行う。本型構造解析部24で処理を行う際、処理中のHTML文書のURIを構造記憶部26に登録する。また、HTML構文解析部23により渡されたリンクを構造記憶部26に存在しないURIが記述されていた場合には、HTML文書取得部22にそれらのURIを渡してHTML文書を取得することを再帰的に行う。取得していないHTM

L文書がなくなったら、本実施形態例で定めた属性に従ってHTML文書の並べ替えを行い、HTML文書の並び方の順番を構造記憶部26に登録し、本型整形部27の処理を開始する。本型整形部27は、構造記憶部26に登録されたHTML文書の並び方の順番で、部品記憶部25に格納されたHTML文書をページに収まるよう分割する処理を行う。本型整形部27で処理を行う際、そのHTML文書のURIとページ番号の対応を記述したURI ⇔ ページ番号対応表28を作成する。本型整形部27の処理が終了したら、その結果を表示データ生成部29に渡す。表示データ生成部29では、本型整形部27で分割されたHTML文書を一ページ毎に情報表示部30で表示できる形式に変換する。表示データ生成部29で処理を行う際、URI ⇔ ページ番号対応表28に存在しないURIは本外部への参照としてそのまま残し、URI ⇔ ページ番号対応表28に存在するURIは本内部への参照としてページ番号に変換する処理を行う。情報表示部30では、表示データ生成部29で変換されたHTML文書を本の形で表示する。

【0034】以上のようにして、HTML文書にそれらの間の論理的構造を記述し、その情報を用いてHTML文書を並べ替えることが可能となり、HTML文書を効率的に本の形に整形し表示することができるようになる。

【0035】次に、図6のフローチャートを参照し、上記実施形態例において本型構造解析部24で本型の論理的構造に基づいたHTML文書の並べ替えを行う動作について詳細に説明する。まず、本型に整形するための出発点となるHTML文書をルート文書と呼ぶ。ルート文書は図1に示した本1に対応し、論理的構造を記述するリンクは本実施形態例で定めた属性bookをもつ。本実施形態例では、最初にルート文書を入力することで処理が開始される。本型構造解析部24では、ステップS1としてHTML構文解析部23より渡されたルート文書のURIを構造記憶部26に格納し、同時に渡されたルート文書内のbook属性をもつリンクに格納されたURIをルート文書に出現する順を並べ、図7に示すような木を作成する。図7の楕円はURIを表しノードと呼ぶ。同図の矢印はbook属性をもつリンクとその属性値である“index”, “section”などを表している。図7のような木を作成するとき、book属性として適さない値をもつリンクは木に含めないとする。次に、現在、葉となっているURIをHTML文書取得部22に渡し、ステップS2に進む。

【0036】ステップS2では、まず、HTML構文解析部23より渡された本実施形態例で定めた属性をもつリンクのうち属性値が“section”, “next”, “previous”であるものを、処理中のHTML文書に出現する順で木に追加する。追加先のノードはHTML構文解析部23より渡されたURIと同じ

URIをもつノードとする。book属性やsection属性の値が“section”であるリンクをsectionリンクと呼ぶが、追加したリンクがsectionリンクだった場合には、その参照先のURIをHTML文書取得部22に渡す。木の中に処理していないsectionリンクが無くなったらステップS3に進む。ステップS2の結果は図8のようになる。同図に示すレベルは、ルート文書から何回のsectionリンク参照で到達できる文書かで定義し、小さい方をより上位のレベルとする。レベル1は図1における章8、レベル2は節9というような対応関係がある。ここで、木へリンクを追加する場合には、それぞれの属性として適さない値をもつリンクや同一もしくはそれ以上のレベルに対するsectionリンクを無視する。

【0037】ステップS3では、木に存在するnextリンク、previousリンクの参照先URIで木の中に存在しないものをHTML文書取得部22に渡し、HTML構文解析部23により渡された本実施形態例で定めた属性をもつリンクのうち属性値“section”, “next”, “previous”であるものを、処理中のHTML文書に出現する順で木に追加する。ここで、nextリンクとは本実施形態例で定めた属性の値が“next”であるリンクのことであり、previousリンクとは本実施形態例で定めた属性の値が“previous”であるリンクのことである。追加先のノードはHTML構文解析部23より渡されたURIと同じURIをもつノードとする。木の中に処理していないnextリンクやpreviousリンクが無くなったらステップS4に進む。

【0038】ステップS4では、木の中に未解決のリンクが含まれているかどうか判定する。未解決のリンクとは、sectionリンク、nextリンク、previousリンクで、その参照先のURIがHTML文書取得部22に渡されていないリンクのことである。未解決のリンクが存在する場合にはステップS2に戻り、未解決のリンクが存在しない場合にはステップS5に進む。

【0039】ステップS5では、同一レベルにあるノードのnextリンク優先の並べ替えを行う。並べ替えは、同一レベルにあるノードをpreviousリンクにできるだけ矛盾がないように並べ替えた後、nextリンクにできるだけ矛盾がないように並べ替えることで行う。ここで、矛盾がないように並べ替えるには、previousリンクやnextリンクによる関係を値の大小関係と考えソートを実行すればよい。

【0040】ステップS5の動作例を、図9を参照しながら説明する。ステップS5の初期状態では図9(a)に示すように、ノードは1番、2番、3番、4番、5番の順で並び、1番から2番にnextリンク、2番から3番にpreviousリンク、3番から4番にnext

tリンク、4番から5番にnextリンクとpreviousリンクが設定されていたとする。previousリンクにできるだけ矛盾がないように並べ替えるには、2番と3番を入れ替え、4番と5番も入れ替えれば良い。すると図10(b)に示すように、1番、3番、2番、5番、4番の順でノードが並ぶことになる。次に、nextリンクにできるだけ矛盾がないように並べ替えるには、4番と5番を入れ替えれば良い。結果、図9(c)に示すように、1番、3番、2番、4番、5番の順でノードが並ぶ。このとき、previousリンクに矛盾が生ずるが、nextリンクが優先であるので無視する。このような並べ替えを木に存在する全てのノードに対して行い、ステップS6に進む。

【0041】ステップS6では、作成された木に従って各ノードに対応するURIを順序づけする。図10は、ステップS6の実行結果の例を示した図である。ステップS5までで、図10(a)のように作成された木を深さ優先で一次元化することで、図10(b)に示したように本型の順序づけができる。また、同図に示したように、木に同一のURIをもつノードが複数存在した場合には、本型に整形されたときに後ろに来るノードを削除する。

【0042】以上で、本発明の第一の実施形態における本型構造解析部24で本型の論理的構造に基づいたHTML文書の並べ替えを行う動作が完了する。

【0043】〔実施形態例2〕次に、本発明の第二の実施形態について図面を用いて詳細に説明する。本実施形態例は、図6に示した本発明の第一の実施形態例に係るHTML文書本型整形装置の構成を表すブロック図における本型構造解析部24を、第一の実施形態例のように本型の論理的構造の記述を全てのHTML文書から引き出すのではなく、本型の論理的構造を記述した目次文書から、その記述を用いて表現されたHTML文書間の論理的構造を本構造に変換するように変更した本発明の一実施形態例である。

【0044】図11は、第二の実施形態におけるHTML文書間の関係を表した図である。角丸四角形はHTML文書内でリンクの記述されている部分を示し、矢印がリンクの参照先を示している。ここで、網掛けした角丸四角形は本の外部への参照を表し、矢印に付加された文字はリンクのbook属性やsection属性、index属性の値を表している。同図に示したmokujiji.html31は、book.html32から、属性book="index"をもつリンクで参照されており、本実施形態例ではここにHTML文書の並び順などの論理的構造が記述される。記述の方法としては、属性index="refer"をもつリンクを本に整形したときの順で記述することが挙げられる。このリンクに、HTML文書内の構造を記述するためのタグである<Hn>タグを組み合わせることで、例えば<H1>

と</H1>で囲まれたリンクは章を表し、<H2>と</H2>で囲まれたリンクは節を表すといったように、HTML文書間の階層関係も表現することができる。

【0045】以下に上記した目次文書であるmokujiji.html31の例を示す。

【0046】mokujiji.html~33

```
<body>
<h1>目次<h1><p>
<h1><A index="refer" href="chap1.html">第一章</A></h1>
<h2><A index="refer" href="sec1.html">第一節</A><h2>
<h2><A index="refer" href="sec2.html">第二節</A><h2>
<p>
```

```
<A index="refer" href="chap2.html">第二章</A>
</body>
```

図12は、図11に示したHTML文書間の関係を用いて本発明の第二の実施形態例によって本型に整形した例を示した図である。それぞれのHTML文書はmokujiji.html33に記述された本型の論理的構造に沿って、図1で示したモデルの本1、目次3、本文4の順に並べられている。本実施形態例でも、一つのHTML文書が一ページに収まりきらない場合は、二ページ以上に分割する。図12に示した各四角形は本型整形後の一ページを表しており、下部にハイフン(-)で囲んだ数字はそのページのページ番号を表している。また、同図の矢印は上記に示した属性index="refer"をもつリンクを表し、点線はHTML文書に記述されているnextリンク、previousリンクによって結びつけられているグループを表している。

【0047】次に、本発明の第二の実施形態例に係るHTML文書本型整形装置の構成を表すブロック図であるが、これは図5に示したものと同様で、HTML構文解析部23から本型構造解析部24へ渡されるデータと本型構造解析部24の動作のみが異なる。HTML構文解析部23の変更は、本型構造解析部24へ渡すデータの内、本実施形態例で定めた属性をもつリンクに、見出しの文字サイズを表す<Hn>タグの現在値を付加するよう変更することで行う。

【0048】本型構造解析部24の動作については第二の実施形態における最も主要な部分であるため、図13のフローチャートを参照しながら詳細に説明する。本実施形態例でも第一の実施形態例と同様に、最初にルート文書を入力することで処理が開始される。本実施形態例の本型構造解析部24では、ステップS7としてHTML構文解析部23より渡されたルート文書のURIを構造記憶部26に格納し、同時に渡されたルート文書内

の属性 `book = "index"` をもつリンクに格納された URI を保持しつつ HTML 文書取得部 22 に渡し、該保持した URI と同じ URI が HTML 構文解析部 23 により渡されるのを待つ。該 URI が HTML 構文解析部 23 より渡されたら、同時に渡された本実施形態例で定めた属性をもつリンクのうち属性 `index = "refer"` であるのを、処理中の HTML 文書に出現する順で木に追加する。追加先のノードは、付加された `<Hn>` タグの値によって決定する。

【0049】図 16 は、第二の実施形態例において目次文書から本型の論理的構造を木に追加している様子を示した図である。目次文書 33 からは、第一章 34、第一章第一節 35、第一章第二節 36、第二章 37 へと属性 `index = "refer"` をもつリンクが設定されているので、そのリンクが出現する順に木への追加を行う。まず、第一章 34 へのリンクは `<H1>` タグと `</H1>` タグで囲まれているのでリンク先はレベル 1 のノードとなる。そこで、第一章 34 へのリンクをルート文書からのリンクとして追加する。次に、第一章第一節 35 へのリンクは `<H2>` タグと `</H2>` タグで囲まれているのでリンク先はレベル 2 のノードとなる。また、最後に追加したレベル 1 のノードは第一章 34 なので、第一章第一節 35 へのリンクを第一章 34 からのリンクとして追加する。同様に、第一章第二節 36 へのリンクも第一章 34 からのリンクとして追加する。最後に、第二章 37 へのリンクには `<Hn>` タグの情報が存在しないため、ルート文書からのリンクとして追加を行う。以上のようにして、目次文書 33 から本の論理的構造を表す木が作成されたら、木の中の未解決のリンクに対してそのリンク先の URI を HTML 文書取得部 22 に渡し、HTML 構文解析部 23 より渡された本実施形態例で定めた属性をもつリンクのうち属性値が `"next"`, `"previous"` であるものを、同時に渡される URI と同じ URI をもつノードに追加する。木の中の未解決のリンクがなくなったら、ステップ S8 に進む。

【0050】ステップ S8 では、ステップ S3 とほとんど同じ動作を行い、木に存在する `next` リンク、`previous` リンクの参照先 URI で木の中に存在しないものを HTML 文書取得部 22 に渡し、HTML 構文解析部 23 より渡された本実施形態例で定めた属性をもつリンクのうち属性値 `"next"`, `"previous"` であるものを、処理中の HTML 文書に出現する順で木に追加する。追加先のノードは同時に渡された URI と同じ URI をもつノードとする。木の中に処理していない `next` リンクや `previous` リンクが無くなったらステップ S9 に進む。

【0051】ステップ S9 では、ステップ S5 と全く同じ動作を行い、同一レベルにあるノードの `next` リンク優先の並べ替えを行う。並べ替えを木に存在する全て

のノードに対して行ったら、ステップ S10 に進む。

【0052】ステップ S10 では、ステップ S6 と全く同じ動作を行い、作成された木に従って各ノードに対応する URI を順序づけし、木に同一の URI をもつノードが複数存在した場合には、本型に整形されたときに後ろに来るノードを削除する。

【0053】以上で、本発明の第二の実施形態例における本型構造解析部 24 で本型の論理的構造に基づいた HTML 文書の並べ替えを行う動作が完了する。

【0054】〔実施形態例 3〕最後に、本発明の第三の実施形態例について詳細に説明する。本実施形態例では、図 6 に示した本発明の第一の実施形態例に係る HTML 文書本型整形装置の構成を表すブロック図における HTML 構文解析部 23 の動作のみが異なる。HTML 構文解析部 23 では、HTML 文書取得部 22 より渡された HTML 文書の構文を解析し、処理中の HTML 文書の URI と本実施形態例で定めた属性をもつリンクを本型構造解析部 24 へ渡し、処理中の HTML 文書を部品記憶部 25 へと格納するが、本実施形態例ではリンクを本型構造解析部 24 へ渡す前に HTML のリンクに従来から存在する REL 属性、REV 属性を本実施形態例で定めた属性に変換して渡すことができる。以下、この変換について詳細に説明する。

【0055】HTML のリンクに従来から存在する REL 属性は、RELATION の略でリンク元からリンク先への順方向の関係を記述する。また、REV 属性は、REVERSE の略でリンク先からリンク元へという逆方向の関係を記述する。REL 属性や REV 属性の値としては、`"made"`, `"parent"`, `"next"`, `"previous"` などが記述できる。そこで、本実施形態例における HTML 構文解析部 23 では、属性 `REV = "parent"` をルート文書では属性 `book = "section"`、その他の HTML 文書では属性 `section = "section"` に変換する。また、属性値が `"next"` や `"previous"` である REL 属性をルート文書では `book` 属性、その他の HTML 文書では `section` 属性に変換する。同様に、ルート文書における属性 `REV = "made"` を属性 `book = "made"` に変換する。このような変換によって、HTML のリンクの属性に対する拡張を行うことなく、HTML 文書間の本型の論理的構造を記述することが可能となる。

【0056】以上で、本発明の第三の実施形態例における HTML 構文解析部 23 で HTML のリンクに従来から存在する REL 属性、REV 属性を本実施形態例で定めた属性に変換する動作が完了する。

【0057】

【発明の効果】以上説明したように、本発明によれば、HTML 文書間に本のような目次、章、節などの論理的構造を付与しようとした場合、その論理的構造に対応し

た属性を付与することで利用者の認識に頼らないリンクを設定することができるようになる。また、上記したようにして記述されたリンクは、他のリンクと属性の点で区別されており、本の形に整形する際にどのリンクを使って順序づけすれば良いかという情報を計算機で抽出することが簡単になるという効果がある。

【0058】さらに、本発明によれば、H T M L 文書間の本型でない論理的構造の記述も損なうことなく本の形に整形し表示することができるため、WWWクライアントと呼ばれる装置に代わって、密な関係をもったH T M L 文書群を本の形で管理する事が可能となり、利用者にとってよりわかりやすい利用法を提供できるようになるという利点もある。

【図面の簡単な説明】

【図1】本発明の第一の実施形態例によって整形された結果の本型の論理的構造をモデル化した図である。

【図2】本発明の第一の実施形態例における論理的構造をリンクに記述するために必要となる属性を示した図である。

【図3】本発明の第一の実施形態例におけるH T M L 文書間の論理的構造を図に表したものである。

【図4】H T M L 文書を本発明の第一の実施形態例によって本型に整形した例を示した図である。

【図5】本発明の第一の実施形態例に係るH T M L 文書本型整形装置の構成を表すブロック図である。

【図6】本発明の第一の実施形態例における本型構造解析部で本型の論理的構造に基づいたH T M L 文書の並べ替えを行う動作を示すフローチャートである。

【図7】図6に示したステップS 1を実行した結果、作成される木の例を示した図である。

【図8】図6に示したステップS 2を実行した結果、図8の木より作成される木の例を示した図である。

【図9】図7に示したステップS 5の動作例を示した図である。

【図10】図7に示したステップS 6の実行結果の例を示した図である。

【図11】本発明の第二の実施形態例におけるH T M L 文書間の関係を表した図である。

*

* 【図12】図11に示したH T M L 文書間の関係を用いて本発明の第二の実施形態例によって本型に整形した例を示した図である。

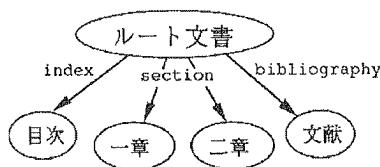
【図13】本発明の第二の実施形態例における本型構造解析部で本型の論理的構造に基づいたH T M L 文書の並べ替えを行う動作を示すフローチャートである。

【図14】本発明の第二の実施形態例において目次文書から本型の論理的構造を木に追加している様子を示した図である。

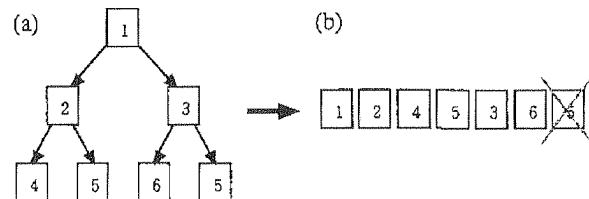
10 【符号の説明】

- 1 …本型の論理的構造モデルにおける本
- 2 …本型の論理的構造モデルにおけるまえがき
- 3 …本型の論理的構造モデルにおける目次
- 4 …本型の論理的構造モデルにおける本文
- 5 …本型の論理的構造モデルにおける参考文献目録
- 6 …本型の論理的構造モデルにおける索引
- 7 …本型の論理的構造モデルにおけるその他の内容
- 8 …本型の論理的構造モデルにおける本文中の章
- 9 …本型の論理的構造モデルにおける章中の節
- 10 …本型の論理的構造モデルにおける節中のページ
- 11 …本型の論理的構造モデルにおけるページ中の単語
- 22 …H T M L 文書取得部
- 23 …H T M L 構文解析部
- 24 …本型構造解析部
- 25 …部品記憶部
- 26 …構造記憶部
- 27 …本型整形部
- 28 …U R I ⇄ ページ番号対応表
- 29 …表示データ生成部
- 30 …情報表示部
- 31 …m o k u j i . h t m l
- 32 …b o o k . h t m l
- 33 …目次文書 m o k u j i . h t m l
- 34 …第一章
- 35 …第一章第一節
- 36 …第一章第二節
- 37 …第二章

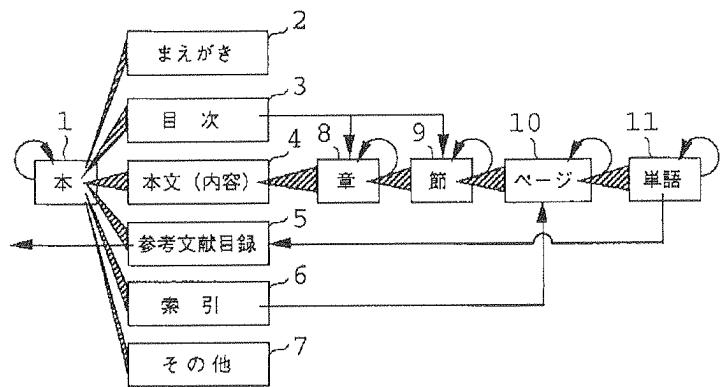
【図7】



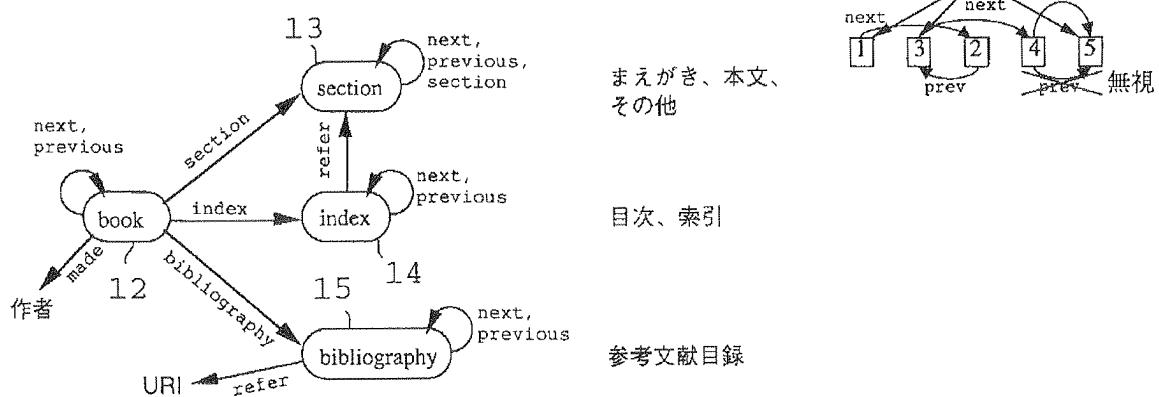
【図10】



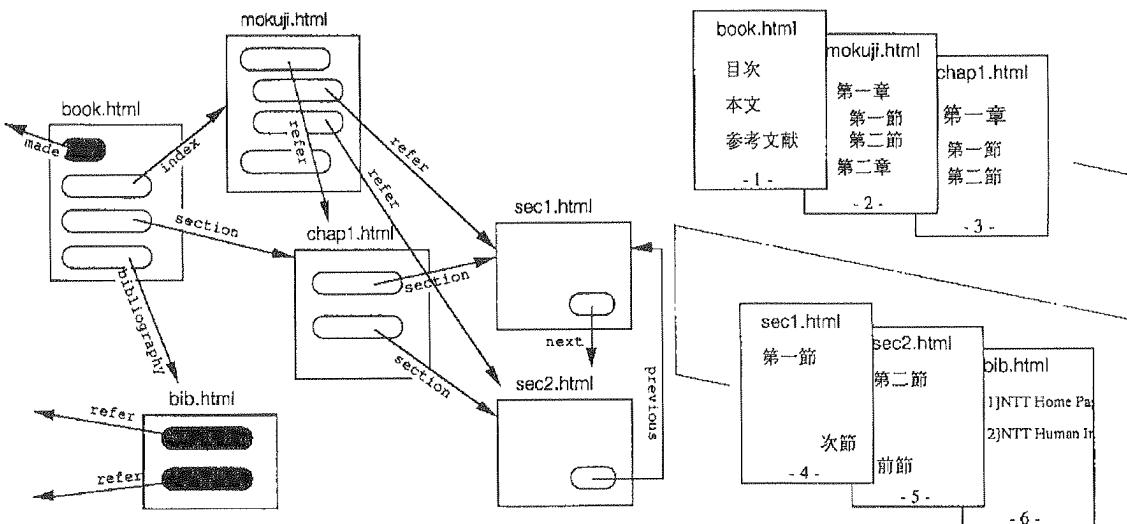
〔四一〕



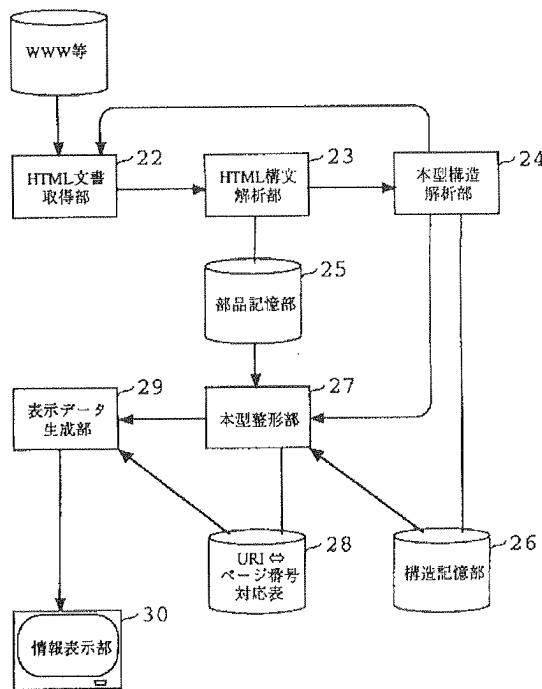
【図2】



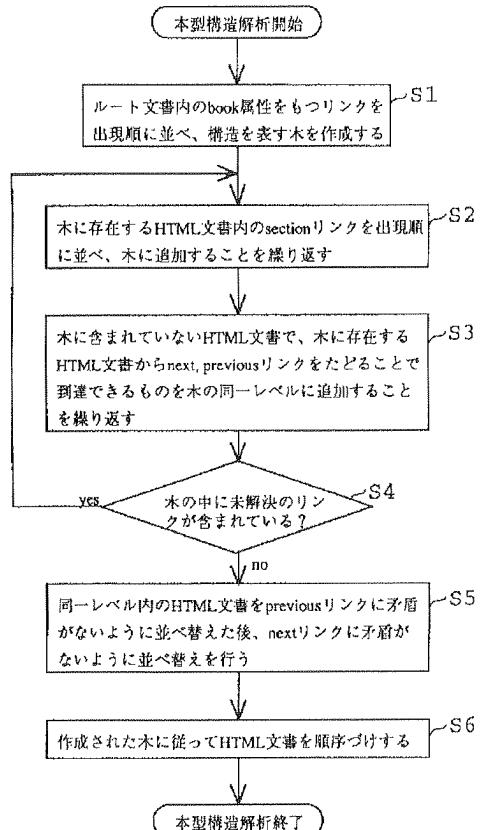
【図3】



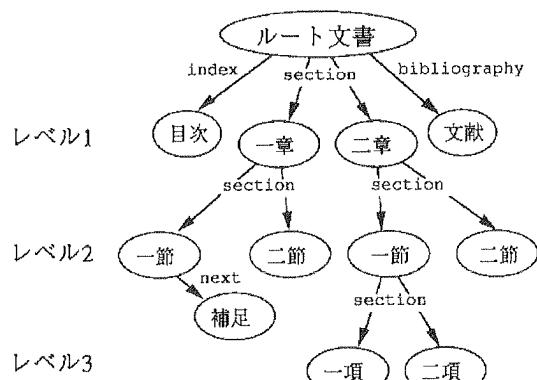
【図5】



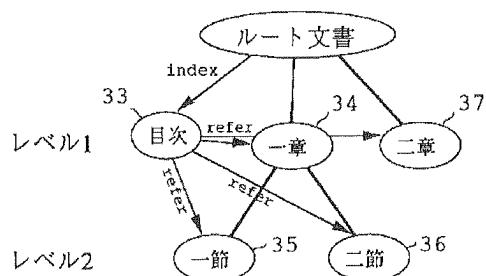
【図6】



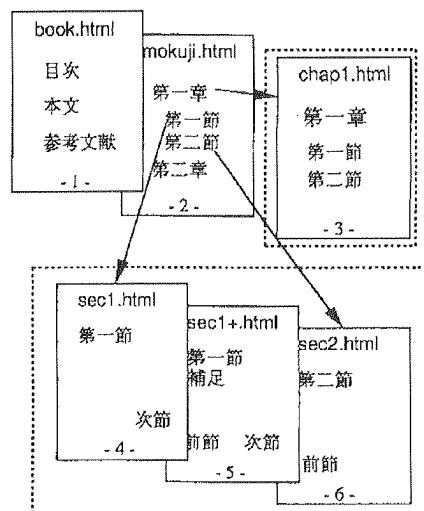
【図8】



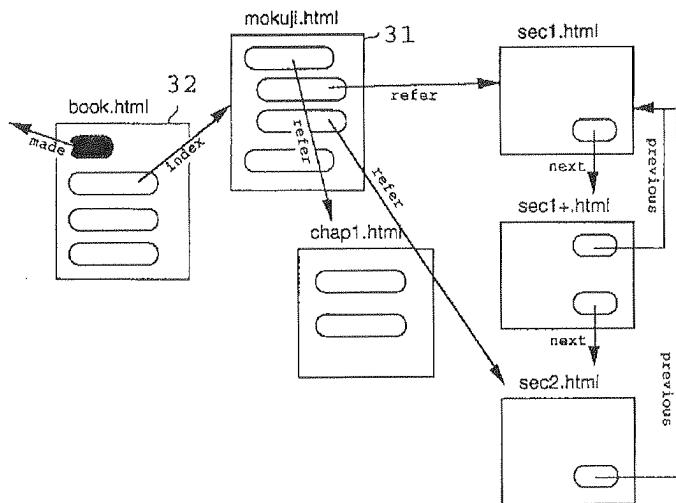
【図14】



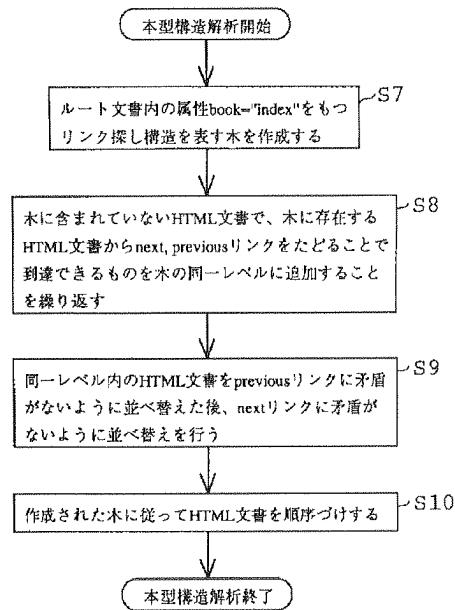
【図12】



【図11】



【図13】



フロントページの続き

(72)発明者 浜田 洋
 東京都新宿区西新宿3丁目19番2号 日本
 電信電話株式会社内